International Journal of Electrical and Electronics Engineering Research (IJEEER)
ISSN(P): 2250-155X; ISSN(E): 2278-943X
Vol. 3, Issue 5, Dec 2013, 145-156

TRANS STELLAR
•Journal Publications •Research Consultancy

COMPARATIVE ANALYSIS OF SPEECH EMOTION RECOGNITION SYSTEM USING DIFFERENT CLASSIFIERS ON BERLIN EMOTIONAL SPEECH DATABASE

RAHUL B. LANJEWAR[1] & D. S. CHAUDHARI[2]

[1]Research Scholar, Department of Electronics and Telecommunication, Government College of Engineering, Amravati, Maharashtra, India

[2]Head, Department of Electronics and Telecommunication, Government College of Engineering, Amravati, Maharashtra, India

ABSTRACT

The man-machine relation has demanded the smart trends that machines have to react after considering the human emotion levels. The technology boost improved the machine intelligence to identify human emotions at expected level. Harnessing the approaches of speech processing and pattern recognition algorithms a smart and emotions oriented man-machine interaction can be achieved with the tremendous scope in the field of automated home as well as commercial applications. This paper deals with the aspects of pitch, Mel Frequency Cepstrum Coefficients based speech features and wavelet domain in speech emotion recognition. The impact of incorporating different classifier using Gaussian Mixture Model (GMM), K-Nearest Neighbour (K-NN) and Hidden Markov Model (HMM) on the recognition rate in the identification of six emotional categories namely happy, angry, neutral, surprised, fearful and sad from Berlin Emotional Speech Database (BES) is emphasized with intents to do a comparative performance analysis.

In the experiments the speech features used are based on pitch, MFCCs and discrete wavelet domain 'db1' family based vectors. The features were same for all the three classifiers of GMM, K-NN and HMM in order to check their comparative performance based on the merits of 'recognition accuracy', 'confusion matrix', 'precision rate' and 'F-measure'. The highest recognition accuracy for the GMM classifiers were 92% for 'angry' emotions, the K-NN classifiers gave 90% correct recognition for 'happy' class, while the highest recognition scores for the HMM classifier were 78% for 'angry' emotions. The confusion matrix statistics depicts the confusion in recognition between 'happy-neutral' emotions; however the three classifiers confused atleast once with the 'angry' emotion in detection of each of the remaining emotions. The results for precision rate and F-measure convey the superiority of GMM classifiers in emotion recognition system while the K-NN and HMM were average in overall performance.

KEYWORDS: Features, Emotion, MFCC, Wavelet, Pitch, K-NN, GMM, HMM, Database

INTRODUCTION

The dynamic requirements of automated systems have pushed the extent of recognition system to consider the precise way of command rather to run only on simple command templates. The idea correlates itself with the speaker identification at the same time recognizing the emotions of speaker. The acoustic processing field not only can identify 'who' the speaker is but also tell 'how' it is spoken to achieve the maximum natural interaction. This can also be used in the spoken dialogue system e.g. at call centre applications where the support staff can handle the conversation in a more adjusting manner if the emotion of the caller is identified earlier.

The human instinct recognizes emotions by observing both psycho-visual appearances and voice. Machines may not exactly emulate this natural tendency as it is but still they are not behind to replicate this human ability if speech processing is employed. Earlier investigations on speech open the doors to exploit the acoustic properties that deal with the emotions. At the other hand the signal processing tools like MATLAB and pattern recognition researcher's community developed the variety of algorithms (e.g. GMM, K-NN, HMM) which completes needed resources to achieve the goal of recognizing emotions from speech [Rahul Lanjewar, 2013] [Ki-Seung Lee, 2008]. This paper focuses on technical challenges that arise when equipping human-computer interface to recognize the user vocal emotions based on the comparative performance of the classifiers of Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) and the K nearest neighbour (K-NN) on Berlin Emotional Speech Database (BES).

RELATED RESEARCH AND MOTIVATION

Busso *et al.* analysed the salient aspects of pitch in which they compared the pitch contour of neutral speech with emotional speech and found that sentence level pitch features has more robustness than voiced-level statistics [Busso *et al.*, 2009]. In different domain approach by Farooq *et al.* analysed wavelet packet transform's multi-resolution capabilities to derived a new superior features than MFCC feature sets and showed improvement in emotion recognition of unvoiced phonemes and stop statements [Faaroq *et al., 2001*]. The Bionic Wavelet transform (BWT) developed by Yao *et.al.* is a biomedical based approach that provides concentrated energy distribution to retain more energy and introduces active control mechanism in auditory system to adjust the wavelet transform. BWT has high sensitivity and selectivity [Yao *et al.,* 2001]. Koolagudi *et.al* reviewed and concluded that entire speech region may not be necessary helps to recognize the underlying emotions. Linear Prediction Cepstrum Coefficients (LPCC), Mel Frequency Cepstrum Coefficients (MFCC) and formants represents the vocal tract information. MFCC are claimed to be robust of all the features for any speech tasks. Wide work in recognition system gave a large variety of classifiers to map the data [Koolagudi *et al.,* 2012]. When the low-level descriptors were employed for classification the HMM gave higher accuracy up to 80% for the speaker dependent recognition [Navas *et al.*, 2006]. Wu *et al.* used new approach of multiple classifiers in which they used classifier models of Gaussian Mixture Model (GMM) [Wu *et al.*, 2011]. But the results of GMM are tending to be more discriminative that of HMM for Berlin Emotional Speech Database (BES) and it measure up to 76% than of 71% of HMM, 67% of K-NN and 55% of FFNN [El Hayadi et *al.*, 2007]. A good and natural database can boost the results of recognition. The speaker specific information always plays an important role. The usage of same speaker for the training and testing the data makes the model to lack generality. Thus the databases with reasonably large speaker and text prompts can give discriminative results at the same time natural emotions database can outclass the simulated database for the real life challenges [Krishnamoorthy *et al.*, 2006]. In a different master class review performed by Ververidis *et al.* on the available 32 database they derived three important results [Ververidis *et al., 2006*]. First, not more than 50% classification can be achieved for the four basic emotions in automated emotion recognition system. Second, simulated emotions are easy to classify as compared to natural emotions. Third, the results usual of emotion recognition in the descending order of their easier classification are anger, sadness, happiness, fear, disgust, joy, surprise and boredom [Kotropoulos *et al.*, 2006].The final purpose of implementing an emotion recognition system is to apply emotion-related knowledge in such a way that human computer communication will be enhanced in such a way that user's experience will become more satisfying.

The rest of the paper is organized as follows: Starting with the system development, the speech features, extraction and selection based on the emotional relevance as identified by the earlier studies is employed and the

comparative analysis of the three classifiers GMM, HMM and K-NN is discussed with respect to their performance merits of 'recognition accuracy', 'confusion matrix', 'precision' and 'F-measure'.

SPEECH EMOTION RECOGNITION SYSTEM

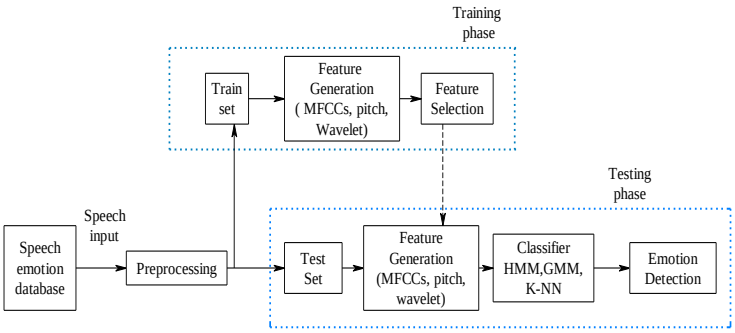

Figure 1: Speech Emotion Recognition System

This section analyzes the design and basic operational aspects of the emotion recognition. Speech emotion recognition is synonymous as a pattern recognition task. This can be understood by the modular flow of process involved in the speech emotion recognition which is shown in the following Figure 1. This sequence is called the pattern recognition cycle. Initially the speech signals are pre-processed. Then they are cut into overlapping frames where the feature extraction algorithms elaborate. The three types of features based on pitch, MFCCs and Wavelet domain were extracted. Afterward the distribution of the feature values with respect to each emotion category is limited by the feature selection process where the 22 MFCC coefficients were selected to reduce the curse of dimensionality while the other features were retained as it is. The selected features were fed to the classifiers. In our study we used three types of classifiers in the form of Gaussian Mixture Model (GMM), K-Nearest Neighbour (K-NN) and Hidden Markov Model (HMM) which is discussed in more details in the following sections. The following sections flourish the work in detail for each of the modules in the speech emotion recognition.

Berlin Emotion Speech Database (BES)

The Berlin emotion speech database is probably the most often used database in the context of emotion recognition from speech, and also one of the few for which some results can be compared. It is one of the databases with acted emotional content. It contains audio recordings of ten actors, five male and five female.

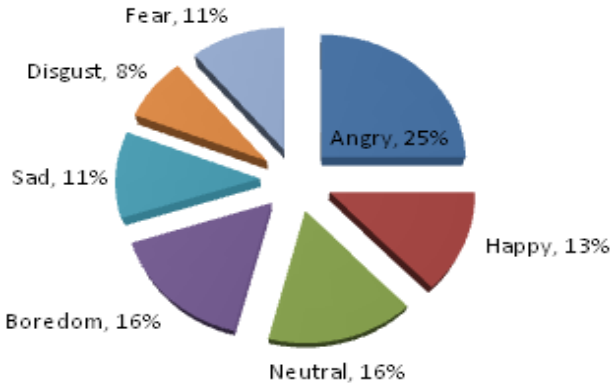

Figure 2: Amount of Recordings from Each Emotion for the Berlin Database

The actors had to portray emotions from the following set: anger, disgust, fear happiness, sadness, surprise and neutral as shown in Figure 2. In order to facilitate their ability to perform naturally, they were asked to induce themselves a

specific state by remembering events that have caused them such emotions. A total of approximately 800 sentences were recorded. After a perception test carried out by 20 participants, the amount was reduced to around 500 samples. The selected utterances have a human recognition rate better than 80% and naturalness scores of more than 60%.

Features Extraction

Any emotion from the speaker's speech is represented by the large number of parameters which is contained in the speech and the changes in these parameters will result in corresponding change in emotions. Therefore an extraction of these speech features which represents emotions is an important factor in speech emotion recognition system [Tawari *et al.*, 2010]. The speech features can be divide into two main categories that is long term and short term features. The region of analysis of the speech signal used for the feature extraction is an important issue which is to be considering in the feature extraction. The speech signal is divided into the small intervals which are referred as frame. The prosodic features are known as the primary indicator of the speakers emotional states. Research on emotion of speech indicates that pitch; energy, duration, formant, Mel frequency cepstrum coefficient (MFCC), and linear prediction cepstrum coefficient (LPCC) are the important features

MFCC Features

A Mel is a unit of measure of perceived pitch or frequency of a tone. The Mel-scale is therefore a mapping between the real frequency scale (Hz) and the perceived frequency scale (mels). The mapping is virtually linear below 1 KHz and logarithmic above. Extensive research on MFCCs indicate that they are less sensitive to noise compared to other currently used parameters and provide better recognition/ identification/ performance than other parameterization schemes. Although the triangular filter bank is used in this study, other windows such as Hamming or Hanning type couldn't be used. After windowing the incoming emotional speech signal, the Discrete Fourier Transform of the frame speech signal is taken. A magnitude spectrum is computed and frequency warped in order to transform the spectrum into Mel frequency in which the filter bank is uniformly spaced [Rabiner, 2005]. The filters multiplied with the magnitude spectra are taken to find the MFCCs as shown in Figure 3. In this study 20 filter banks and 22 MFCCs are used for the simulations.

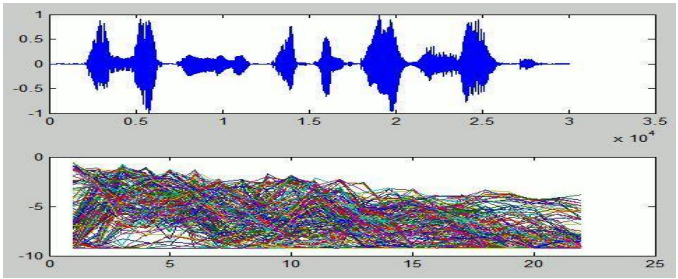

Figure 3: Mel Power Cepstrum for 'Angry' Speech

The WAVELET Features

The detail discussion on wavelet analysis is beyond the scope of this paper and the more complete discussion is presented in [Sarikaya *et al.*, 1998]. The continuous wavelet transform is defined here. Let $f(t)$ be any square integrable function. The CWT or continuous-time wavelet transform of $f(t)$ with respect to a wavelet $\psi(t)$ is defined as:

$$W(a,b) \equiv \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a}\right) dt$$

Where a and $b$are real and * denotes complex conjugation. Thus the wavelet transform is a function of two variables. Both the $f(t)$ and $\psi(t)$ belongs to the set of energy signals. The equation can be written in compact form by defining $\psi_{a,b}(t)$ as:

$$\psi_{a,b}(t) \equiv \frac{1}{\sqrt{|a|}}\psi^*\left(\frac{t-b}{a}\right)$$

Wavelet functions comprise an infinite set. The different wavelet families make different trade-offs between how compactly the basis functions are localized in space and how smooth they are. The Haar wavelet is discontinuous and resembles a step function. It represents the same wavelet as Daubechies db1. In this paper we considered db1 family of wavelets for the speech features extraction as shown in Figure 4.

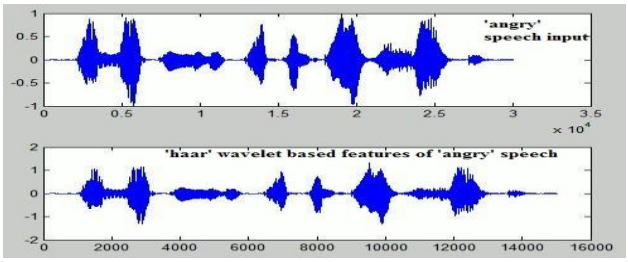

Figure 4: Waveform of 'Angry' Speech and its 'Haar' Wavelet Feature

Pitch Features

The pitch features are derived using a pitch determination algorithm based on Subharmonic-to-Harmonic Ratio (SHR). The magnitude of subharmonics with respect to harmonics reflects the degree of deviation from modal voice. The SHR reflects the ratio of amplitudes of harmonics and subharmonics. The algorithm is based on the pitch perception study to determine the perceived speech and SHR. The technique involves synthesis of vowels with alternate cycles through amplitude and frequency modulation, which generated subharmonics with lowest frequency of $0.5F_0$. Generally, when the ratio is smaller than 0.2, the subharmonics do not have effects on pitch perception. As the ratio increases approximately above 0.4, the pitch is mostly perceived as one octave lower that corresponds to the lowest subharmonic frequency. When SHR is between 0.2 and 0.4, the pitch seems to be ambiguous. These findings suggest that pitch could be determined by computing SHR and comparing it with the pitch perception data. The procedure for computing SHR falls in the general category of spectrum compression technique [Xuejing Sun, 1999]. A sample of speech and its pitch is shown in Figure 5.

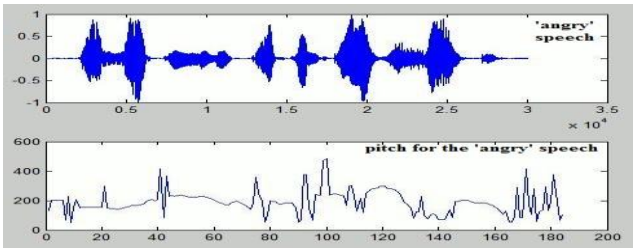

Figure 5: Pitch (f0) Features for 'Angry' Speech

CLASSIFICATION

In the speech emotion recognition system after calculation of the features, the best features are provided to the classifier. A classifier recognizes the emotion in the speaker's speech utterance. Nowadays, research is focused on finding

powerful combinations of classifiers that advance the classification efficiency in real life applications. Various types of classifiers have been proposed for the task of speech emotion recognition which is given as below.

The Gaussian Mixture Model (GMM)

In this study, a Gaussian Mixture Model approach is proposed where speech emotions are modelled as a mixture of Gaussian densities. The use of this model is motivated by the interpretation that the Gaussian components represent some general emotion dependent spectral shapes and the capability of Gaussian mixtures to model arbitrary densities.

The Gaussian Mixture Model is a linear combination of M Gaussian densities, and given by the equation,

$P(\vec{x}|\lambda)= \sum_{i=1}^{M} p_i \, b_i(\vec{x})$

where $\vec{x}$is a D-dimensional random vector, $b_i(\vec{x})$, i=1,.....,M are the component densities and p_i, i=1,.....M are the mixture weights. Each component density is a D-dimensional Gaussian function of the form

$b_i(\vec{x})=\frac{1}{(2\pi)^2|\sum i|^2} \exp\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \sum_i^{-1}(\vec{x}-\vec{\mu}_i)\}$

where$\vec{\mu}_i$ denotes the mean vector and $\sum_i$ denotes the covariance matrix. The mixture weights satisfy the law of total probability, $\sum_{i=1}^{M} p_i$=1. The major advantage of this representation of speaker models is the mathematical tractability where the complete Gaussian mixtures density is represented by the mean vectors, covariance matrices and mixture weights from all component densities [El Ayadi *et al.*, 2007].

K-Nearest Neighbour (K-NN)

A general version of the nearest neighbour technique bases the classification of an unknown sample on the "votes" of K of its nearest neighbour rather than on only it's on single nearest neighbour. If the costs of error are equal for each class, the estimated class of an unknown sample is chosen to be the class that is most commonly represented in the collection of its K nearest neighbours [Pao *et al.*, 2008].

Let the *k* neighbours nearest to *y* be N_k(Y) and c(*z*) be the class label of *z*. The cardinality of N_k(Y) is equal to *k* and the number of classes is *l*. Then the subset of nearest neighbours within class $j \in \{1, ..., l\}$ is:

N_k^j(Y)= {Z∈ N_k(Y)*: c(z)=j*}

$j^* \in \{1, \ldots . l\}$

The classification result {1,..., } * $j \in l$ is defined as the majority vote: j^*=arg$max_j \left|N_k^j(Y)\right|$

Hidden Markov Model (HMM)

Hidden Markov models can be regarded as the simplest dynamic Bayesian networks (DBN). They have a long tradition in speech recognition based on the idea that the statistics of voice are not stationary. The use of HMM and their capability to model the temporal behaviour of speech as opposed to the global statistics approach has more advantages. The HMM consists of the first order Marko chain whose states are hidden from the observer therefore the internal behaviour of the model remains hidden. The hidden states of the model capture the temporal structure of the data. Hidden Markov models are the statistical models that describe the sequences of events. HMM is having the advantage that the temporal dynamics of the speech features can be trapped due to the presence of the state transition matrix. During classification, a speech signal is taken and the probability for each speech signal is provided to the model is

calculated. An output of the classifier is based on the maximum probability that the model has been generated this signal [Nogueiras *et al.*, 2001] [Ntalampiras *et al.*, 2012].

EXPERIMENTAL RESULTS AND ANALYSIS

Detecting emotions of happy, fear, angry, neutral, sad and surprise from the speech signal was the main motive of this work. Of all the system modules the database played a vital role. In the evaluation of the system, apart from the Berlin Emotion database a new database is recorded to check the detection accuracy of the standard as well as of weak database. The new database is recorded contains 40 utterance of word "Good Morning" by 8 different speakers, 5 male and 3 female. The main purpose of this idea was to extend the work and validate the techniques of our proposed system results and the results of the previous work in which HMM and SVM were used as a classifiers and MFCC as the speech features [Chaudhari *et al.*, 2012]. In our approach we used HMM, GMM and K-NN as classifiers with Wavelet features as well as MFCC and pitch features of speech are employed. Thus from the available resources of speech features (MFCC, wavelet, pitch), availability of database (BES) and classifiers (GMM, HMM, K-NN) a comparative as well as cross validation approach is adopted to obtain the recognition results in the form of recognition accuracy, confusion matrix, precision rate and F-measure on BES for all the three classifiers. The same three classifiers are used on recorded non-standard database as a purpose of comparative analysis since the previous work consist only MFCC features as speech features and in our work we used an extra feature sets based on pitch and Wavelet analysis of input speech.

Recognition Accuracy

This measure signifies the recognition accuracy in percentage for each known test speech input to the total trained emotional speech data.

$$Accuracy = \frac{Correctly\ detected\ Emotions\ inputs}{Total\ trained\ emotions\ inputs} \text{ X } 100\%$$

The accuracy for each classifier for the six emotions is calculated on the basis of above relation. It is calculated forthe Berlin Emotional Speech Database (BES) as shown in Figure 6

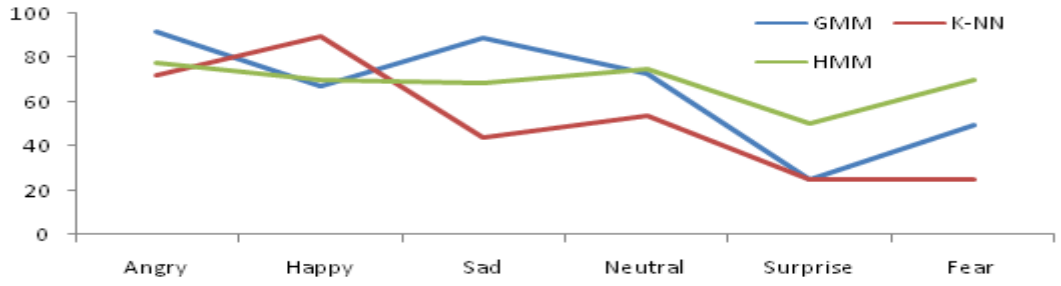

Figure 6: Recognition Accuracy

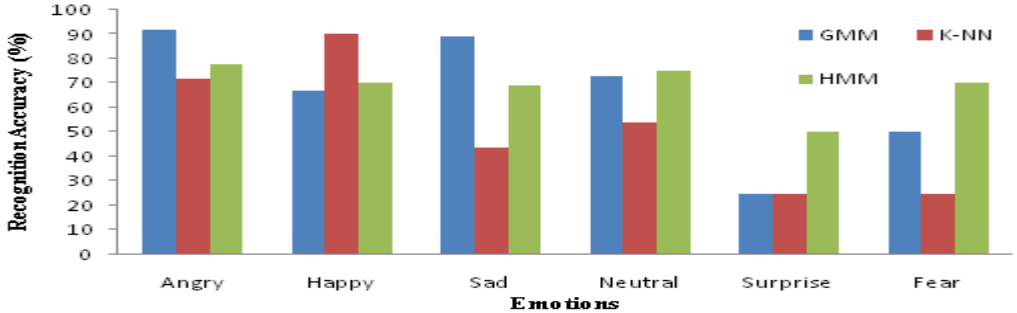

Figure 7

Confusion Matrix

The confusion matrix of the speech emotion recognition system signifies the slackness of classifiers to choose the best correct emotion in the testing phase. It depicts the confusion in selection among the trained patterns of emotion features having similarities in their respective feature pattern shown in Table1, Table 2 and Table 3.

Table.1: Confusion Matrix for K-NN Classifier

Responded / Presented	Angry	Happy	Sad	Neutral	Surprise	Fear
Angry	72	-	-	28	-	-
Happy	-	90	-	10	-	-
Sad	12	22	44	-	-	22
Neutral	28	-	-	54	4	14
Surprise	50	-	-	-	25	25
Fear	41	-	25	-	-	34

Table 2: Confusion Matrix for GMM Classifier

Responded / Presented	Angry	Happy	Sad	Neutral	Surprise	Fear
Angry	92	-	-	-	8	-
Happy	-	67	-	-	33	-
Sad	-	-	89	11	-	-
Neutral	-	19	-	73	-	8
Surprise	25	25	-	-	25	25
Fear	-	-	-	-	50	50

Table 3: Confusion Matrix for HMM Classifier

Responded / Presented	Angry	Happy	Sad	Neutral	Surprise	Fear
Angry	65	-	-	5	-	30
Happy	-	80	-	20	-	-
Sad	26	12	64	-	-	-
Neutral	-	20	-	78	2	-
Surprise	10	20	-	-	70	-
Fear	20	-	20	-	10	50

Precision Rate

It is defined as the ratio of correctly recognized emotions for each class to the correctly recognized emotions for all the classes [17].

$$Precision\ Rate\ (\%) = \frac{Correctly\ recognized\ emotions\ for\ each\ class}{Correctly\ recognized\ emotions\ for\ all\ the\ classes}$$

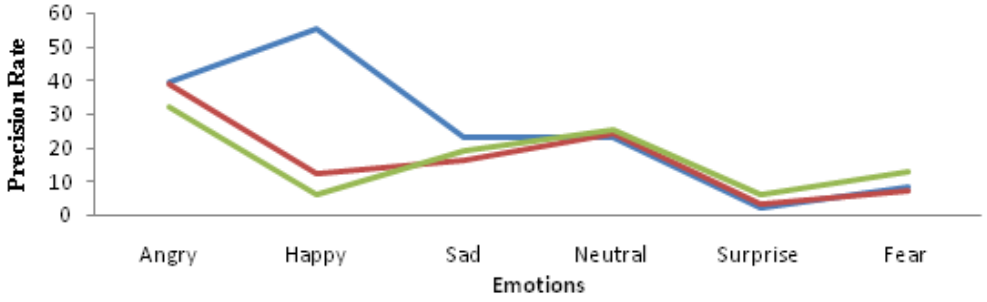

Figure 8: Precision Rates (%) for Each Class of Emotions under Different Classifiers

The results of precision rates are lowest for the 'surprise' emotion class while they are nearby similar for 'neutral' emotions. The 'angry' class of emotion has highest precision rates of recognition for all the three classifiers. The GMM classifier has shown tremendous precision rate for all class of emotions while the K-NN has average rates but the HMM has shown lowest performance as compare to other classifiers as shown in Figure 8

F-Measure

The F-Measure is the merit of combination of precision rate and accuracy. It is adopted from the work of Natalampiras *et al*. in which the performance of the implementation was evaluated from this factor to obtain the overall performance of the system in term s of correct results *i.e.* by not considering the wrong recognition observations and is given by [17]:

$$F - Measure = \frac{2*Accuracy*Precision}{Accuracy+Precision}$$

The F-Measure of the proposed system as shown in Figure 9 for the three classifiers displays the overall performance of the system in terms of its correctness.

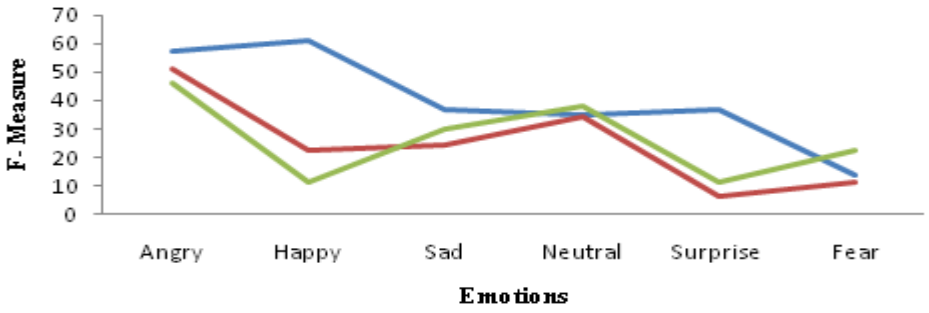

Figure 9: F-Measure of the System for the GMM, K-NN and HMM

This merit depicts the superiority of GMM classifier for all the emotions as compared to K-NN and HMM classifier. The K-NN performance has slumped as compared to its initial performance at the same time the HMM technique has improved for the tailing emotions of the graph shown in Figure 9

DISCUSSIONS AND CONCLUSIONS

The purpose of incorporating different classifiers to recognize the emotions for the same types of speech features helps us draw a comparative conclusion based on their performance. In classification the GMM technique has shown best results in detecting all the emotions with minimum recognition rate of 25% for 'surprise' emotion to the highest rate of 92% for 'angry' emotion. The confusion matrix supports the recognition results of GMM since maximum confusion occurred in detecting between 'surprise-happy' emotions. The recognition rates for the K-NN technique are also promising with their statistics in detecting 'happy' emotion with 90% rate and lowest rate for "fear' while for the 'surprise' emotion it was 50%. The confusion matrix show maximum confusion to recognize among 'fear-angry' and 'sad-angry' emotions for the K-NN technique. The HMM technique is overall average in recognizing "angry' emotion with 78% while the maximum confusion was with 'angry-surprise' but as compared to other two techniques used the HMM technique has shown sank results. These results are experimented on BES database while the results for the recorded speech database are dropped because of very poor performance of system which gave fatal error sometimes as a consequence of employing poor quality of data which lacked the emotional features and thus resulted in the performance of recognition for all the three classifiers.

However the confusion matrix results of K-NN technique to recognize between happy and neutral emotions adds similar relevance to natural techniques since there are circumstances in which even humans get confuse to judge between 'happy' and 'normal' emotions. The HMM and GMM techniques erred in recognizing 'angry' emotion among the rest of the five emotions. Last but not least if a robust and efficient emotion recognition technique is to be implemented on real time applications then the three techniques when fused together can recognize the emotions very effectively because of their dominance for particular type of emotions as such GMM for 'angry' and 'sad' emotions detection, K-NN for 'happy' as well as 'angry' emotions and HMM for the 'fear' and 'normal' emotions detection. However, the merits of Precision rate and F-Measure has depicted the overall performance of the system in terms of the all the correct recognition results in which the GMM technique emerged as a robust system of the proposed speech emotion recognition system. Last but not least if a robust and efficient emotion recognition technique is to be implemented on real time applications then the three techniques when fused together can recognize the emotions very effectively because of their dominance for particular type of emotions as such GMM for 'angry' and 'sad' emotions detection, K-NN for 'happy' as well as 'angry' emotions and HMM for the 'fear' and 'normal' emotions detection. The speed of computation was less for K-NN classifier makes it as one of the optional techniques that can be used widely if time constraint does not matters. The time of computation increased for GMM classifier when the number of speech features increased in training phase. However the HMM classifier took maximum time to detect the emotions, its computation time increased drastically as soon as the feature set increased which make it less potent in hardware implementation for time constraint applications.

REFERENCES

1. Ashish Tawari and Mohan Manubhai Trivedi, 'Speech Emotion Analysis: Exploring the Role of Context', *IEEE Trans. on Multimedia*, Vol. 12, No. 6, 2010, pp 502-509.
2. Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, and José B. Mariño, 'Speech Emotion Recognition Using Hidden Markov Models', Eurospeech 2001 – Scandinavia.
3. Chang-Hsein Wu and Wei-Bin Liang, "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic and Semantics Labels, *IEEE Trans. on Affective Computing*, Vol 2, No.1, 2011, pp.567-569.
4. C. Busso, S. Lee and S. Narayanan, "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection", *IEEE Trans. on Audio, Speech and Language processing*, Vol. 17, No. 4,2009, pp 582-596.
5. D.S.Chaudhari, Ashish B. Ingale, "Speech Emotion Recognition Using Hidden Markov Model and Support Vector Machine", Int'l Journal of Advanced Engineering Research and Studies, Vol. 1, 2012, pp.316-318.
6. Dimitrious Ververidis and Constantine Kotropoulos, "A Review of Emotional Speech Databases", Speech Communication, 2006, pp.1162-1181.
7. Eva Navas, Inmaculada Hernáez, and Iker Luengo, 'An Objective and Subjective Study of the Role of Semantics and Prosodic Features in Building Corpora for Emotional TTS', *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 14, No. 4, July 2006, pp. 1117-1127.
8. Jun Yao and Yuan-Ting Zhang, 'Bionic Wavelet Transform: A New Time–Frequency Method Based on an Auditory Model', *IEEE Trans*. on *Biomedical Engineering*, Vol. 48, No. 8, 2001, pp.856-863.
9. Ki-Seung Lee, "EMG-Based Speech Recognition Using Hidden Markov Models With Global Control Variables", *IEEE Trans. on Biomedical Engineering*, Vol. 55, No. 3, 2008, pp.136-138.

10. Moataz M. H. El Ayadi, Mohamed S. Kamel and Fakhri Karray, "Speech Emotion Recognition using Gaussian Mixture Vector Autoregressive Models', ICASSP, 2007, pp.957-960.
11. O. Farooq and S. Datta, 'Mel Filter-Like Admissible Wavelet Packet Structure for Speech Recognition', *IEEE Signal Processing Letters*, Vol. 8, No. 7, 2001, pp.196-198.
12. P.Krishnamoorthy and S.R. Mahadeva Prasanna, "Temporal and Spectral Processing methods for Processing of Degrading Speech: A Review", IETE Technical Review,Vol 26, Issue 2, Apr 2009, pp. 87-90.
13. Rabiner L. R.and Juang B, "Fundamentals of Speech Recognition", Pearson Education Press, Singapore, 2nd edition, 2005
14. Rahul. B. Lanjewar, D. S. Chaudhari, "Speech Emotion Recognition: A Review", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 2, March 2013, pp. 68-71.
15. Ruhi Sarikaya, Brian L. Pellom and John H.L. Hansen, 'Wavelet Packet Transform Features with Application to Speaker Recognition', 1998, pp.912-915.
16. Shashidhar G. Koolagudi ·K. Sreenivasa Rao, 'Emotion recognition from speech: a review', Int'l Journal on Speech Technology, 2012, pp.99–117.
17. Stavros Ntalampiras and Nikos Fakotakis,' Modeling the Temporal Evolution of Acoustic Parameters for Speech Emotion Recognition', *IEEE Trans. on Affective Computing*, Vol. 3, No. 1, 2012, pp. 116-125.
18. Tsang-Long Pao, Wen-Yuan Liao and Yu-Te Chen, 'A Weighted Discrete KNN Method for Mandarin Speech and Emotion Recognition', Speech Recognition Technologies and Applications, 2008, pp. 550-552.
19. Xuejing Sun, 'Pitch Determination and Voice Quality Analysis using Subharmonic-To-Harmonic Ratio', Department of Communication Sciences and Disorders, Northwestern University, 1999, pp. 561-563.